

Hammar, Olle

Working Paper

A Comment on "Income and Inequality in the Aztec Empire on the Eve of the Spanish Conquest"

I4R Discussion Paper Series, No. 221

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Hammar, Olle (2025) : A Comment on "Income and Inequality in the Aztec Empire on the Eve of the Spanish Conquest", I4R Discussion Paper Series, No. 221, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/315048>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



No. 221

I4R DISCUSSION PAPER SERIES

A Comment on “Income and Inequality in the Aztec Empire on the Eve of the Spanish Conquest”

Olle Hammar

April 2025

I4R DISCUSSION PAPER SERIES

I4R DP No. 221

A Comment on “Income and Inequality in the Aztec Empire on the Eve of the Spanish Conquest”

Olle Hammar¹

¹Linnaeus University, Växjö/Sweden

APRIL 2025

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

A comment on “Income and inequality in the Aztec Empire on the eve of the Spanish conquest”

Olle Hammar
(Linnaeus University and Institute for Futures Studies)*

Abstract

Alfani and Carballo (2023) estimate the levels of income inequality in the Aztec Empire around 1492, that is, before the Spanish conquest. Their main estimate finds that the Gini index was 50.4. They conclude that income inequality in the Aztec Empire was high even before the Spanish conquest, questioning to what extent today's high levels of economic inequality in Mexico can be explained by the Spanish conquest and extractive institutions imposed by the colonizers. First, I confirm that the main outcomes are computationally reproducible from the analysis data provided in the replication package. Second, I detect two inconsistencies with respect to the Kolmogorov-Smirnov test and the IV analysis, but after correcting for those I obtain qualitatively similar results. Third, I test the robustness reproducibility of the relationship between population density and per capita income through eleven different robustness tests, confirming a positive and statistically significant relationship but with large variation in the point estimates. Finally, I use these estimates as inputs for robustness measures of the main outcomes. On average, these robustness reproductions yield very similar results to those in the original paper. However, they also indicate large uncertainty about the exact estimates, for example, with Gini index estimates ranging between 38.8 and 65.6. As such, the conclusion that the level of income inequality was higher in the Aztec Empire than in modern Mexico does not appear to be a robust finding. The finding that it was more unequal than in contemporary United States, however, seems to be robust.

* Email: olle.hammar@lnu.se.

1. Introduction

Alfani and Carballo (2023) estimate the levels of income inequality and imperial extraction in the Aztec Empire around 1492, that is, before the Spanish conquest. Their analysis consists of two steps. First, they estimate the relationship between population density and per capita income in preindustrial societies. Second, they use these regression results as inputs for calculations of their main outcomes, namely the levels of income and inequality in the Aztec Empire. Their main estimates find per capita income levels of 690 USD (1990 PPP), corresponding to 1.72 times the subsistence minimum. In terms of income inequality, they find that the richest 1% earned 41.8% of total income, the richest 5% earned 50.8% of total income, and the poorest 50% earned 23.3% of total income. Their estimates suggest that the imperial ruling class extracted 4.54% of total income, and that the Gini index was 50.4. They conclude that income inequality in the Aztec Empire was high even before the Spanish conquest, questioning to what extent today's high levels of economic inequality in Mexico can be explained by the Spanish conquest and extractive institutions imposed by the colonizers.

In the present report, prepared as part of a collaboration between the Institute for Replication (I4R) and *Nature Human Behaviour* (Brodeur et al. 2024), I investigate whether their analytical results are computationally reproducible and further test their replicability and robustness to: (1) including country and year fixed effects, (2) including all observations, (3) using absolute values, (4) using the full Ashraf and Galor (2011) sample, (5) using population density from Ashraf and Galor (2011), and (6) using “plants” as instrument. When possible, the robustness analyses are done for both models used in the original paper (OP), that is, ordinary least squares (OLS) and instrumental variables (IV). As such, in total, I conduct eleven robustness reproducibility analyses. Finally, I compare these robustness reproduction results to those in the OP, including their estimated income distribution in the Aztec Empire, and those of modern Mexico and the United States.

Overall, the estimates in the OP seem robust in terms of averages. At the same time, however, the large dispersion of estimates from my robustness analyses suggest that they are surrounded by large uncertainties.

2. Computational Reproducibility

For the computational reproducibility, I used the original authors' replication package available here: <https://www.openicpsr.org/openicpsr/project/186521/version/V2/view>.

The replication package included two parts. In the first part, the analysis data and codes for the econometric regression analyses were provided. In this part, the per capita income levels were estimated. These estimates were then used as inputs for the second part. The first part was conducted in Stata.

In the second part, the income estimates from the first part, together with other input data, were used to estimate the levels of income inequality. These analyses were done in Microsoft Excel.

No raw data nor cleaning codes were provided. Analysis code was missing for the Kolmogorov-Smirnov tests comparing the estimated income distribution of the Aztec Empire to those of modern Mexico and the USA (OP Figure 2 and p. 1266).

All main outcomes were computationally reproducible from the analysis data.

Table 1 – Results of computational reproduction

	Fully	Partial	No
Raw data provided			x
Cleaning code provided			x
Analysis data provided	x		
Analysis code provided		x	
Reproducible from raw data			x
Reproducible from analysis data	x		

As an alternative to the original authors' Excel calculations, I also computationally reproduced the main Gini index (for the Aztec Empire as a whole) using the Stata command *ineqdeco* (Jenkins 1999). This reproduced the original Gini estimate of 50.4.

To reproduce the Kolmogorov-Smirnov tests in the note to OP Figure 2, for which no analysis code was provided, I used the Stata command *ksmirnov*. This yielded the same or very similar D-values, but larger p-values (see Table 2). In the OP, the differences between the income distributions in the Aztec Empire compared with modern American states were reported as statistically significant at the 1% level. In the reproduction, however, the income distribution difference between the Aztec Empire and modern Mexico was only significant at the 10% level, and the distribution difference between the Aztec Empire and the USA was significant at the 5% level.

Table 2 – Computational reproduction results for Kolmogorov-Smirnov tests

Test for equality of income distributions	Reported in OP (Figure 2)	Reproduction using <i>ksmirnov</i>
Aztec Empire 1492 versus Mexico 2018	D = 0.55*** (p-value = 0.004)	D = 0.55* (exact p-value = 0.075)
Aztec Empire 1492 versus USA 2018	D = 0.65*** (p-value = 0.000)	D = 0.64** (exact p-value = 0.021)

Note: Significant at the ***[1%] **[5%] *[10%] level.

A coding error was found in the main IV analysis (Model 2), which used the *reg* and *predict* commands in Stata. In this analysis, the restriction of the natural logarithm of population density having to be greater than or equal to zero was only imposed in the first stage and not in the second stage. Moreover, estimating the IV model in two steps, instead of using a single command such as *ivregress*, resulted in different standard errors. The Hausman specification test for choosing between the OLS Model 1 and the two-stage least squares (2SLS) Model 2 (see discussion in OP p. 1269), however, used the *ivreg* command (but without robust standard errors). The different results are presented in Table 3 (Columns 1–3).

Table 3 – Computational reproduction results for original IV analysis

Model 2 (IV)	Original used in main analysis (1)	Original using same samples in both stages (2)	Original used in Hausman test using <i>ivreg</i> (3)	Reproduction original corrected using <i>ivregress</i> (4)
(Ln)	0.198***	0.157***	0.157	0.157***
Population density	(0.048) [0.000]	(0.048) [0.002]	(0.095) [0.104]	(0.054) [0.005]
Constant	5.829*** (0.151) [0.000]	5.807*** (0.234) [0.000]	5.807*** (0.304) [0.000]	5.807*** (0.313) [0.000]
Controls	Yes	Yes	Yes	Yes
Observations	66	57	57	57
R-squared	0.086	0.048	0.275	0.275

Note: IV regressions using 2SLS. Number of domesticable species of *animals* used as instrumental variable. The dependent variable is the natural logarithm of income per capita. The main independent variable is the natural logarithm of population density. Control variables include natural logarithm of land productivity, natural logarithm of absolute latitude, mean distance to nearest coast or river, and percentage of land within 100 km of coast or river. Population density from Maddison (2003). Small-sample degrees-of-freedom adjustments, except for (1) and (2). Adjusted R-squared, except for (1) and (2). Robust standard errors are in parentheses for (1), (2) and (4), and non-robust standard errors are in parentheses for (3). P-values are reported in brackets. Significant at the ***[1%] **[5%] *[10%] level.

Table 3 (Column 4) also reports the corresponding results using the *ivregress* command (with robust standard errors and small-sample degrees-of-freedom adjustments), which is what I have used in the remainder of this reproduction. As shown in Table 3, the coefficient estimates were the same for the corrected reproduction (Column 4) as for the one used in the Hausman test in the OP (Column 3), that is, 0.157, but smaller than the original estimate used in the main analysis and reported in the OP (Column 1), that is, 0.198. While the original results used in the Hausman test without robust standard errors were not statistically significant (Column 3), the corrected results with robust standard errors were found statistically significant at the 1% level (Column 4).

The authors did not register a pre-analysis plan.

3. Robustness Reproduction

3.1 Robustness reproduction of the population density and income relationship

I conducted four pre-specified robustness reproductions of the analysis estimating the relationship between population density and per capita income in preindustrial societies (see pre-re-analysis plan in the Appendix): (1) including country and year fixed effects (FE) instead of other control variables, (2) including all observations including those where the logarithmic population density variables is smaller than zero, (3) using absolute values instead of natural logarithms for population density and per capita income, and (4) using the full sample of observations provided by Ashraf and Galor (2011). These robustness reproduction analyses were pre-specified to focus on the OLS model (Model 1). When possible, I have also conducted them for the IV model (Model 2), which was the main model used in the OP.

In addition to the pre-specified robustness reproductions, I have also conducted two exploratory (that is, non-pre-specified) robustness reproductions, based on reading Ashraf and Galor's (2011) paper and looking at their data: (5) using log population density from Ashraf and Galor (2011) instead of Maddison (2003), and (6) using "plants" instead of "animals" as instrumental variable.

The robustness reproducibility results for Model 1 (OLS) are presented in Table 4, and the robustness reproducibility results for Model 2 (IV) are presented in Table 5.

Table 4 – Robustness reproduction results for OP Model 1 (OLS)

Model 1 (OLS)	Original	Country and year fixed effects	All observations	Absolute values	Full AG2011 sample	Population density from AG2011
	(1)	(2)	(3)	(4)	(5)	(6)
(Ln)	0.189***	0.335***	0.083***	11.037***	0.084***	0.080***
Population	(0.040)	(0.044)	(0.012)	(2.369)	(0.012)	(0.011)
density	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Constant	5.728***	5.456***	5.995***	394.201***	5.999***	6.062***
	(0.387)	(0.094)	(0.165)	(113.253)	(0.164)	(0.162)
	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]
Controls	Yes	No	Yes	Yes	Yes	Yes
Fixed effects	No	Yes	No	No	No	No
Observations	66	66	88	88	89	86
R-squared	0.200	0.847	0.192	0.483	0.192	0.315

Note: The dependent variable is the natural logarithm of income per capita, except for (4) where it is the absolute values of income per capita. The main independent variable is the natural logarithm of population density, except for (4) where it is the absolute values of population density. Control variables include natural logarithm of land productivity, natural logarithm of absolute latitude, mean distance to nearest coast or river, and percentage of land within 100 km of coast or river. Population density from Maddison (2003), except for (6) where it is from Ashraf and Galor (2011) [AG2011]. Robust standard errors are in parentheses. P-values are reported in brackets. Significant at the ***[1%] **[5%] *[10%] level.

Table 5 – Robustness reproduction results for OP Model 2 (IV)

Model 2 (IV)	Original	Original corrected	All observations	Absolute values	Full AG2011 sample	Population density from AG2011	Plants as instrument
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(Ln)	0.198***	0.157***	0.118***	14.706***	0.121***	0.107***	0.350
Population density	(0.048) [0.000]	(0.054) [0.005]	(0.032) [0.000]	(4.620) [0.002]	(0.032) [0.000]	(0.026) [0.000]	(0.233) [0.139]
Constant	5.829*** (0.151) [0.000]	5.807*** (0.313) [0.000]	5.909*** (0.242) [0.000]	320.851** (150.267) [0.037]	5.916*** (0.243) [0.000]	6.033*** (0.197) [0.000]	5.541*** (0.720) [0.000]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	66	57	66	66	67	65	67
R-squared	0.086	0.275	0.253	0.373	0.254	0.269	.
First-stage F	.	13.097	10.996	11.232	11.058	16.258	1.621

Note: IV regressions using 2SLS. Number of domesticable species of *animals* used as instrumental variable, except for (7) where number of domesticable species of *plants* is used instead. The dependent variable is the natural logarithm of income per capita, except for (4) where it is the absolute values of income per capita. The main independent variable is the natural logarithm of population density, except for (4) where it is the absolute values of population density. Control variables include natural logarithm of land productivity, natural logarithm of absolute latitude, mean distance to nearest coast or river, and percentage of land within 100 km of coast or river. Population density from Maddison (2003), except for (6) where it is from Ashraf and Galor (2011) [AG2011]. Small-sample degrees-of-freedom adjustments, except for (1). Adjusted R-squared, except for (1). Columns (1) and (2) same as columns (1) and (4) in Table 3, respectively. Robust F-statistics reported from first-stage regressions. Robust standard errors are in parentheses. P-values are reported in brackets. Significant at the ***[1%] **[5%] *[10%] level.

3.1.1 Country and year fixed effects

First, the results could be affected by the choice of control variables included and excluded in the empirical specification estimating the relationship between population density and per capita income in preindustrial societies. As such, in a first reproducibility robustness check, I included country and year fixed effects instead of other control variables. These results are presented in Table 4 (Column 2). The relationship was still found positive and statistically significant, but with a larger point estimate (0.335) than in the original OLS specification (0.189).

3.1.2 *All observations*

Second, in the original analysis the sample was restricted to the observations where the logarithmic population density was greater than zero. In a second robustness check, I included all observations including those where the logarithmic population density variable was smaller than zero. This increased the sample size from 66 to 88 country-years in the OLS specification (additional countries in the larger sample were Algeria, Australia, Brazil, Canada, Finland, Libya, New Zealand, Norway, and the United States), and from 57 to 66 observations in the IV specification (additional countries Brazil, Finland, and Norway). These results are presented in Table 4 (Column 3) for the OLS results, and in Table 5 (Column 3) for the IV results. The relationship was found positive and statistically significant, but with smaller point estimates (0.083 for OLS and 0.118 for IV) than in the original specifications.

3.1.3 *Absolute values*

Third, as shown in the OP Supplementary Figure 1, there was a positive correlation between population density and per capita income both in absolute values and for their natural logarithmic transformations. As a third robustness check, I thus used absolute values instead of natural logarithms for population density and per capita income. These results are presented in Table 4 (Column 4) for OLS, and in Table 5 (Column 4) for IV. The results were found positive and statistically significant.

3.1.4 *Full Ashraf and Galor (2011) sample*

Fourth, according to the OP, the analysis was conducted using “a sample of 66 pre-industrial societies, which included 22 distinct world areas all observed in three different years (1, 1000 and 1500 AD)” (p. 1269), based on data from Ashraf and Galor (2011). This statement, however, was not fully correct. The original OLS analysis included 66 observations, but not from a balanced panels of countries and years. Instead, it included 24 countries, of which 21 were observed in all three years, while 3 countries were only observed in one year. The first stage in the original IV analysis included 21 countries, of which 18 were observed in all three years. As a fourth robustness check, I used the full sample of observations provided by Ashraf and Galor (2011). This increased the sample

size to 89 country-years in the OLS specification and to 67 country-years in the IV specification (the additional country relative to the “All observations” robustness specification in Column 3 was Tunisia, observed in year 1). These larger-sample results are presented in Table 4 (Column 5) for OLS and in Table 5 (Column 5) for IV. The relationship between population density and per capita income was still found positive and statistically significant, but with smaller coefficients (0.084 for OLS and 0.121 for IV) than found in the original specifications.

3.1.5 Population density from Ashraf and Galor (2011)

Fifth, the data from Ashraf and Galor (2011) included two different estimates of the log population density: one compiled from various historical sources (Ashraf and Galor 2011), and one taken from Maddison (2003). In the OP, only the latter variable, that is, based on Maddison (2003), was used. Hence, as a fifth (non-pre-specified) robustness check, I instead used Ashraf and Galor’s (2011) main log population density variable (that is, the one based on various sources). The results are presented in Table 4 (Column 6) for OLS, and in Table 5 (Column 6) for IV. The relationship was still found positive and statistically significant, but with smaller point estimates (0.080 for OLS and 0.107 for IV) than in the OP.

3.1.6 Plants as instrument

Sixth, and finally, Ashraf and Galor (2011) provided two potential instruments for the IV specification, namely the number of domesticable species of 1) plants and 2) animals that were prehistorically native to the continent or landmass to which a country belongs. In the OP, only the latter (that is, *animals*) was used as an instrumental variable. As a sixth (non-pre-specified) robustness check, I have instead used the number of domesticable species of *plants* as instrumental variable in the IV specification. These alternative IV results are presented in Table 5 (Column 7). The coefficient estimate was positive, but not statistically significant. However, the first-stage results indicate that “plants” did not satisfy the relevance assumption and, as such, was not a valid instrument.

3.2 Robustness reproduction of the main outcomes: income and inequality

Overall, the robustness reproduction results in Section 3.1 showed that the positive and statistically significant relationship between population density and per capita income was robust and reproducible. This result, however, was not a main finding of the OP. Instead, the coefficients from these regressions were used as input variables for estimating the main outcomes: income levels, income inequality, and imperial inequality in the Aztec Empire. Next, I have thus used the coefficient estimates from Section 3.1 as inputs to analyze the robustness reproducibility of the following main outcomes: absolute per capita income, relative per capita income, Gini index, richest 1% income share, richest 5% income share, poorest 50% income share, and extraction imperial ruling class. The calculations of these outcomes were done in the same way as in the OP, but using the coefficient estimates from the robustness regressions as inputs. The results are presented in Table 6.

Table 6 – Robustness reproduction results for main outcomes

Model	Specification	Per capita income		Income inequality	Income share (%)			Imperial inequality (%)
		Absolute (1990 PPP)	Relative (sub. min.)	Gini index	Richest 1%	Richest 5%	Poorest 50%	Extraction ruling class
IV.1	Original	690	1.72	50.4	41.8	50.8	23.3	4.54
OLS.1	Original	611	1.53	44.2	36.2	44.4	26.3	5.12
OLS.2	Country and year FE	771	1.93	55.6	47.5	56.0	20.9	4.06
OLS.3	All observations	557	1.39	38.8	30.7	39.1	28.9	5.61
OLS.4	Absolute values	907	2.27	62.2	55.6	62.6	17.7	3.45
OLS.5	Full AG2011 sample	560	1.40	39.1	30.9	39.3	28.8	5.59
OLS.6	Pop. density AG2011	586	1.46	41.7	33.2	42.0	27.5	5.34
IV.2	Original corrected	593	1.48	42.5	34.5	42.7	27.1	5.27
IV.3	All observations	575	1.44	40.6	32.5	40.9	28.0	5.45
IV.4	Absolute values	998	2.49	65.6	59.7	66.0	16.1	3.14
IV.5	Full AG2011 sample	583	1.46	41.4	33.3	41.7	27.6	5.37
IV.6	Pop. density AG2011	621	1.55	45.0	36.3	45.3	25.9	5.04
IV.7	Plants as instrument	879	2.20	61.0	52.4	61.4	18.3	3.56
<i>Mean (robustness)</i>		<i>694</i>	<i>1.73</i>	<i>48.5</i>	<i>40.6</i>	<i>48.8</i>	<i>24.3</i>	<i>4.72</i>

Note: Comparison between estimates of income levels, income inequality, and imperial inequality in OP and robustness reproductions. Mean (robustness) measured as average values of OLS.2–6 and IV.2–7. For description of how regression results are used to calculate income and inequality levels, see OP.

3.2.1 *Income levels*

In the OP, the absolute per capita income for the Aztec Empire as a whole was estimated to 690 USD (1990 PPP) using the original IV specification, and to 611 USD using the original OLS specification. The average value for my robustness reproduction estimates was 694 USD, that is, very similar to the OP estimate. At the same time, there is relatively large variation in the values estimated in the different robustness specifications, ranging from 560 USD to 998 USD, suggesting a high degree of uncertainty about the true value. For example, this questions the statement in the OP that “[the estimate] is much lower than the \$800–900 (1990 PPP) of late fifteenth-century Spain” (p. 1266). For relative per capita income, in the OP, this was estimated to 1.72 times the subsistence minimum in the IV specification, and to 1.53 subsistence minimums in the OLS specification. In my different robustness specifications, this value ranged between 1.39 and 2.49, with an average value of 1.73 subsistence minimums.

3.2.2 *Income inequality*

For income inequality, in the OP, the Gini index was estimated to 50.4 using Model 2 (IV) and to 44.2 using Model 1 (OLS). In my robustness estimations, the average value for the Gini index was 48.5, with estimates ranging from 38.8 to 65.6. As a comparison, a Gini index of 65.6 is higher than the value for the country with the highest level of income inequality today (that is, South Africa with a Gini index of 63.0), while a Gini index of 38.8 corresponds to that of the 58th–60th most unequal countries today (Laos, Gambia, and El Salvador) according to the World Bank (2025). The income share of the richest 1% was estimated to 41.8% of total income (36.2% for OLS) in the OP. The average value in my robustness analyses was 40.6%, with estimates ranging between 30.7% and 59.7%. For the income share of the richest 5%, the OP estimate was 50.8% (44.4% for OLS). The average robustness reproduction estimate was 48.8%, with estimated values between 39.1% and 66.0%. The income share of the poorest 50% was estimated to 23.3% (26.3% for OLS) in the OP. My robustness estimates ranged from 16.1% to 28.9%, with the

average estimated income share of the bottom half being 24.3% of total income. As with income levels, the OP income inequality estimate thus seems a plausible estimate, albeit with large uncertainty.

3.2.3 Imperial inequality

The OP also estimated the level of imperial extraction across the Aztec Empire. The imperial ruling class extraction was estimated to 4.54% of total income in Model 2 (IV), and 5.12% in Model 1 (OLS). The average estimate in my robustness analyses was 4.72%, with estimates ranging between 3.14% and 5.61% of total income.

3.3 Testing the statistical difference between the income distributions

Finally, in the OP a Kolmogorov-Smirnov test was conducted to test if the estimated income distribution of the Aztec Empire in 1492 was statistically different from those of modern Mexico and the United States in 2018. The analysis in the OP found that the distribution was statistically different (with a higher level of inequality in the Aztec Empire than in modern Mexico and United States). While I was not able to fully reproduce this result (see Table 2 in Section 2), the computational reproduction qualitatively confirmed this finding (but with larger p-values than in the OP).

In a final analysis, I have used the Aztec Empire income distributions estimated from the different robustness reproductions and tested if they are statistically different from 1) the estimated Aztec Empire distribution in the OP, 2) the income distribution in modern Mexico, and 3) the income distribution in modern United States. The comparisons were conducted through a Kolmogorov-Smirnov test and based on estimated income shares of deciles and the top 5% income group. The results are presented in Table 7.

Table 7 – Kolmogorov-Smirnov tests comparing income distributions

Model	Specification	OP Aztec Empire 1492 (1)	Mexico 2018 (2)	USA 2018 (3)
IV.1	Original	.	0.55* [0.075]	0.64** [0.021]
OLS.1	Original	0.45 [0.211]	0.55* [0.075]	0.55* [0.075]
OLS.2	Country and year fixed effects	0.45 [0.211]	0.64** [0.021]	0.64** [0.021]
OLS.3	All observations	0.55* [0.075]	0.45 [0.211]	0.55* [0.075]
OLS.4	Absolute values	0.82*** [0.001]	0.73*** [0.004]	0.73*** [0.004]
OLS.5	Full AG2011 sample	0.55* [0.075]	0.45 [0.211]	0.55* [0.075]
OLS.6	Population density from AG2011	0.45 [0.211]	0.45 [0.211]	0.55* [0.075]
IV.2	Original corrected	0.45 [0.211]	0.45 [0.211]	0.55* [0.075]
IV.3	All observations	0.45 [0.211]	0.45 [0.211]	0.55* [0.075]
IV.4	Absolute values	0.82*** [0.001]	0.73*** [0.004]	0.73*** [0.004]
IV.5	Full AG2011 sample	0.45 [0.211]	0.45 [0.211]	0.55* [0.075]
IV.6	Population density from AG2011	0.45 [0.211]	0.55* [0.075]	0.55* [0.075]
IV.7	Plants as instrument	0.73*** [0.004]	0.64** [0.021]	0.73*** [0.004]

Note: Results from Kolmogorov-Smirnov tests comparing the estimated income distributions for the Aztec Empire from the different reproduction robustness analyses with (1) the distribution estimated in the OP, (2) the distribution in Mexico 2018, and (3) the distribution in the United States 2018, using the *ksmirnov* command in Stata. Estimations based on income distributions in deciles and the top 5% as a separate group. Combined Kolmogorov–Smirnov D. Exact p-values are reported in brackets. Difference between the distributions significant at the ***[1%] **[5%] *[10%] level.

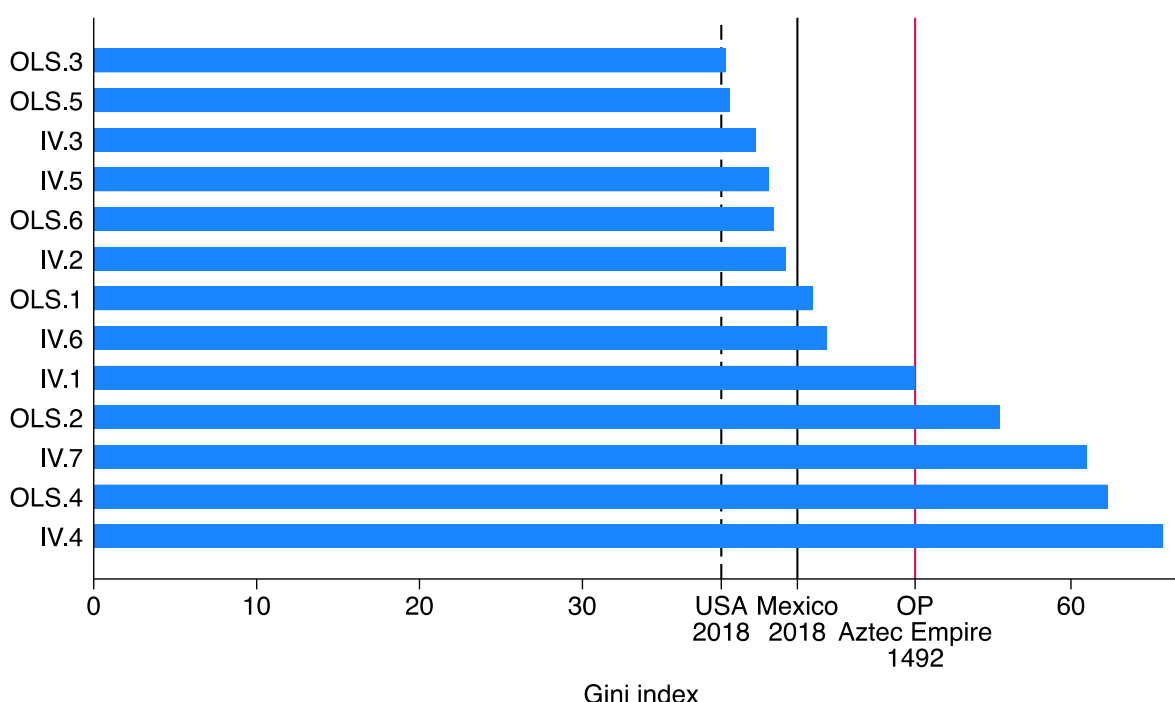
The results from the Kolmogorov-Smirnov tests reported in Table 7 show that out of the eleven different robustness reproduction analyses, six yielded income distributions that were not statistically different from the OP. However, three were statistically different at the 1% significance level, and another two were statistically different at the 10% significance level.

Comparing the estimated distributions from the robustness reproductions with the income distribution in modern Mexico, those were not statistically different in six specifications, statistically different at the 10% level in one specification, and at the 5% or 1% significance

level in four specifications. The comparison with the income distribution in modern United States showed statistically different distributions in all cases (at the 10% significance level in seven specifications, and at the 5% or 1% level in four specifications). In other words, the OP result that incomes in the Aztec Empire was more unequally distributed than those in modern United States seems to be a robust finding. The OP conclusion that “[t]he income distribution of the Aztec Empire differs significantly from those of modern-day Mexico” (p. 1266), however, does not appear to be a robust result.

Finally, a graphical comparison of the different Gini index estimates is illustrated in Figure 1.

Figure 1 – Gini index comparison between OP and robustness reproductions



Note: Gini indices for the Aztec Empire computed based on the coefficient estimates from the different robustness reproduction regressions, as well as the OP's Gini index for the Aztec Empire (red solid line), modern Mexico (black solid line), and the United States (black dashed line). For description of the different specifications, see main text.

Figure 1 shows the relatively large variation in Gini indices estimated for the Aztec Empire in the different robustness reproductions. While, on average, the robustness estimates were similar to the one in the OP, the exact value appears very uncertain. All robustness

estimations resulted in higher Gini coefficients than the one of modern United States, supporting the OP result that income inequality was higher in the Aztec Empire than in contemporary United States. Comparing the income distribution in the Aztec Empire with the one in current Mexico, however, approximately half of the estimations resulted in lower Gini coefficients and half in higher Gini coefficients. Again, the OP finding that income inequality was higher in the Aztec Empire than in modern Mexico does not appear to be a robustly reproducible result.

4. Conclusion

Overall, the findings of Alfani and Carballo (2023) were computationally reproducible. On average, my replicability robustness checks also yielded very similar results to those found in the original analysis. At the same time, the large variation in reproduction estimates emphasized the large uncertainty surrounding the exact estimates of income and inequality in the Aztec Empire. Finally, it should also be noted that, in these robustness analyses, I have only varied the specifications for the regression analysis estimating the relationship between population density and per capita income. Future replications could for instance also analyze robustness with respect to variations in the raw input data and/or the model for estimating the levels of income and inequality.

References

- Alfani, G., and Carballo, A. (2023). Income and inequality in the Aztec Empire on the eve of the Spanish conquest. *Nature Human Behaviour* 7, 1265–1274.
<https://doi.org/10.1038/s41562-023-01636-3>.
- Ashraf, Q., and Galor, O. (2011). Dynamics and stagnation in the Malthusian epoch. *American Economic Review* 101, 2003–2041.
- Brodeur, A., Dreber, A., Hoces de la Guardia, F., and Miguel, E. (2024). Reproduction and replication at scale. *Nature Human Behaviour* 8, 2–3.
<https://doi.org/10.1038/s41562-023-01807-2>.
- Jenkins, S. P. (1999). INEQDECO: Stata module to calculate inequality indices with decomposition by subgroup. *Statistical Software Components* S366002. Boston College Department of Economics.
- Maddison, A. (2003). *The World Economy: Historical Statistics*. Development Centre Studies. Paris: OECD.
- World Bank (2025). Gini index. *DataBank*.
<https://data.worldbank.org/indicator/SI.POV.GINI> (accessed February 7, 2025).

Appendix

I4R NHB Pre-re-analysis plan

Name of replicator: Olle Hammar

Title of paper: Income and inequality in the Aztec Empire on the eve of the Spanish conquest

Original authors: Guido Alfani & Alfonso Carballo

I will conduct a computational and robustness reproduction of Alfani and Carballo (2023) “Income and inequality in the Aztec Empire on the eve of the Spanish conquest” (*Nature Human Behaviour* 7: 1265–1274), including the following steps:

1. The analysis in the original paper is conducted in Stata and Excel. Since codes are only provided for parts of the analysis, a first check will be to recode the main analysis in Stata based on the information provided in the original article and its supplementary materials.
2. Next, I will check the robustness of the estimation results analyzing the relationship between population density and per capita income for preindustrial societies, focusing on the OLS model (Model 1). The robustness checks will consist of four alterations:
 - a. The results could be affected by the choice of control variables included/excluded in the specification. As such, in a first robustness check, I will include country and year fixed effects instead of other control variables.
 - b. In the original analysis, the sample is restricted to the observations where the logarithmic population density is greater than zero. In a second robustness check, I will include all observations including those where the logarithmic population density variable is smaller than zero.
 - c. As shown in the original paper’s Supplementary Figure 1, there is a positive correlation between population density and per capita income both in absolute values and for their natural logarithmic transformation. Hence, as a third robustness check, I will use absolute values instead of natural logarithms for population density and per capita income.
 - d. The original analysis is conducted using “a sample of 66 pre-industrial societies, which included 22 distinct world areas all observed in three different years (1, 1000 and 1500 AD)”, based on data from Ashraf and Galor (2011) “Dynamics and stagnation in the Malthusian epoch” (*American Economic Review* 101: 2003–2041). As a final robustness check, I will use the full sample of observations provided by Ashraf and Galor (2011).

3. Finally, I will check to what extent these robustness analyses change the following main outcomes (measured for the Aztec Empire as a whole):
- a. Income levels
 - i. Per capita income, absolute: **US\$690** (1990 PPP)
 - ii. Per capita income, relative: **1.72** times subsistence minimum
 - b. Income inequality
 - i. Gini index: **50.4**
 - ii. Richest 1% income share: **41.8%** of total income
 - iii. Richest 5% income share: **50.8%** of total income
 - iv. Poorest 50% income share: **23.3%** of total income
 - c. Imperial inequality
 - i. Extraction imperial ruling class: **4.54%** of total income

To test whether the income distribution becomes different from that reported in the paper, I will use the Kolmogorov–Smirnov test.